What is claimed is:

1. A computerized storage and retrieval system of biological information, comprising:

a data entry means;

a display means;

a programable central processing unit; and

a data storage means having cDNA sequences and related information electronically stored in a relational database;

wherein the stored sequences are annotated and organized in a curated functional clustering arrangement.

2. The computerized storage and retrieval system of claim 1, wherein the central processing unit is programmed with the ability to calculate significance values, perform gene annotation analysis, generate transcript images, perform transcript image analysis, perform subtractive analysis, perform electronic Northern analysis, or perform electronic commonality analysis.

3. The computerized storage and retrieval system of claim 1, wherein the central processing unit is programmed to perform automated bioanalysis on the stored cDNA sequences.

4. The computerized storage and retrieval system of claim 3, wherein the automated bioanalysis comprises:

sequence editing; and

annotation and organization of sequences.

5. The computerized storage and retrieval system of claim 4, wherein the sequence editing comprises the steps of:

identifying and clipping vector sequences;

identifying and clipping unwanted functional motifs;

identifying and removing cloning and sequencing artifacts; and

identifying and masking low information sequences.

6. The computerized storage and retrieval system of claim 4, wherein the automated bioanalysis further comprises the steps of:

transcript extension; and

transcript expansion.

7. The computerized storage and retrieval system of claim 1, wherein the stored cDNA sequences are comprised of SEQ ID NOS. 1-10.

8. The computerized storage and retrieval system of claim 1, wherein the information pertaining to the cDNA sequences is stored in a plurality of tables, said tables organized into categories.

9. The computerized storage and retrieval system of claim 8, wherein the categories comprise library preparation, clone preparation, sequencing, sequencing equipment, sequencing reagents, function identification, and express sets.

10. The computerized storage and retrieval system of claim 1, wherein the storage means can be searched to determine source tissue information, to determine source organ information, to determine source pathology information, or to determine source patient information.

11. A method for quantifying the relative abundance of mRNA species in a sample, said method comprising the steps of:

generating cDNA sequences corresponding to a representative population of transcripts found within a sample;

organizing cDNA sequences into a functional clustering arrangement;

accessing a computerized storage and retrieval system of biological information containing reference cDNA sequence data corresponding to full-length reference transcripts annotated and stored in a curated functional clustering arrangement;

processing the sample sequence data and the reference sequence data in a programmed computer to generate an identified sequence value for each of the gene transcripts, said sequence value being indicative of a sequence annotation and a degree of match between a transcript from the sample sequence data and at least one transcript from the reference sequence data; and

processing each identified sequence value to generate final data values indicative of a number of times each identified sample sequence value is present within the curated reference sequences.

12. The method of claim 11, wherein the reference sequence data is comprised of SEQ ID NOS. 1-10.

13. The method of claim 11, wherein the method is used for transcript discovery.

14. A method for transcript image analysis comparing two or more samples, comprising the steps of:

producing a first transcript image from the cDNA sequences corresponding to a representative population of the full-length transcripts of a first sample;

producing a second transcript image from the cDNA sequences corresponding to a representative population of full-length transcripts of a second samples; and

determining the ratio of the frequency that a cDNA sequence appears in the first and the second source tissues;

wherein the ratio of the frequency that a cDNA sequence appears in the first and the second samples is indicative of the relative levels of expression of the corresponding transcripts in each sample.

15. The method of claim 14 wherein said second sample is a stored reference set of cDNA sequences.

16. The method of claim 14, wherein the first sample is from normal tissue and the second sample is from a diseased or potentially diseased sample.

17. The method of claim 14, wherein the method is used for transcript discovery.

18. A method for performing electronic Northerns comprising the steps of:

selecting libraries corresponding to samples of interest;

selecting a cDNA sequence to examine in each selected library; and

performing abundance analysis for said cDNA sequence in each library;

wherein the abundance of cDNA sequence in each library is indicative of the location, distribution, and relative abundance of gene expression in the selected samples.

19. The method of claim 18, wherein the gene expression is determined by the abundance of a cDNA sequence selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

20. The method of claim 18, wherein the method is used for diagnostic purposes, prognostic purposes, or determining patient treatment.

21. A method for performing electronic commonality analysis between a first sample and a second sample, said method comprising the steps of:

producing a first transcript image from the cDNA sequences corresponding to a representative population of the full-length transcripts of a first sample;

producing a second transcript image from the cDNA sequences corresponding to a representative population of full-length transcripts of a second sample; and

electronically comparing the transcript images of the sample data set and the reference data set to identify transcripts expressed in both of the two samples;

wherein normalized abundances are used to determine a ratio of expression between the two samples.

22. The method of claim 21 wherein said second sample is a stored reference set of cDNA sequences.

23. The method of claim 21, wherein the first sample is from normal tissue and the second sample is from a diseased or potentially diseased sample.

24. The method of claim 21, wherein the sample is selected from the group consisting of blood, sputum, urine, ascites fluid, cerebrospinal fluid, and biopsy tissue.

25. The method of claim 21, wherein the method is used for transcript discovery.

26. The method of claim 21, wherein the method is used for diagnostic purposes, prognostic purposes, or determining patient treatment.

27. A method for performing electronic subtraction analysis between a first sample and a second sample, said method comprising the steps of:

producing a first transcript image from the cDNA sequences corresponding to a representative population of the full-length transcripts of a first sample;

producing a second transcript image from the cDNA sequences corresponding to a representative population of full-length transcripts of a second sample ; and

selecting a target abundance value for transcripts found in the sample sequence data and reference sequence data;

processing this information to determine the transcripts within each sequence data set that exceed the selected target abundance value; and

electronically comparing the transcript images of the sample data set and the reference data set to identify transcripts expressed in only one of the two samples.

28. The method of claim 27 wherein said second sample is a stored reference set of cDNA sequences.

29. The method of claim 27, wherein the first sample is from normal tissue and the second sample is from a diseased or potentially diseased sample.

30. The method of claim 27, wherein the sample is selected from the group consisting of blood, urine, sputum, ascites fluid, cerebrospinal fluid, and biopsy tissue.

31. The method of claim 27, wherein the method is used for transcript discovery.

32. The method of claim 27, wherein the method is used for transcript discovery, diagnostic purposes, prognostic purposes, or determining patient treatment.

ADD 7
A2